**Literature Review**

Predicting the result of a court case based on the facts of the case is known as legal judgement prediction.

Previously, only feature-based models (e.g., utilizing a bag of words and subjects) have been explored in English when employing neural networks for this job in Chinese. A new English judicial decision prediction dataset, which includes judgments from the European Court of Human Rights, has been made available by the researchers. On the new dataset, we test a wide range of neural models and create strong baselines that outperform earlier feature-based models in three tasks: (1) binary violation classification; (2) multi-label classification; (3) case significance prediction. Data anonymization allows us to see whether models are skewed toward demographic information. BERT's length constraint may be circumvented using a hierarchical variant of the algorithm.

Textual representations of legal information are common (e.g., legal cases, contracts, bills). There are several applications for analyzing legal texts, including legal subject categorization [1]. [1], court opinion legal information generation and analysis [2].

Legal judgement prediction is what we're focusing on here, where we try to anticipate the result of a legal case by looking at a written description of the facts. With these models, legal practitioners and people may benefit from reduced legal expenses and improved access to justice. Judges can use them to evaluate the chances of winning a case and come to more consistent, accurate decisions. Educated decisions, to put it another way. Human rights groups and academics may make use of them to examine the fairness of court rulings and reveal if prejudices are linked to them [3].

Some researchers use three major task for the configuration (1) binary configuration on human rights articles violation happened, (2) multi-label classification ( it is a type of violation), (3) and another one is case priority detection. But all of this work is done by the use of SVM bag-of-word in neural network [4]

One of previous work contributions is to examine, whether legal prediction models are skewed toward demographic information or human rights-relevant facts using an approach based on anonymized data.

**Datasets used:**

Human rights violations are investigated by the European Court of Human Rights (ECHR). 2 The ECHR's public database comprises around 11.5k cases in dataset. Regular expression extraction of facts from case descriptions is provided in each dataset case, as Altera’s and colleagues (2016).

Articles that were violated in each instance are also linked to each individual case (if any). European Committee for Human Rights (ECHR) assigns a significance score to each case.

A number of subsets of the dataset have been created for various purposes. Cases from 1959 to 2013 are included in the training and development sets, while tests from 2014 to 2018 are included in the test set. There are equal numbers of training and development sets with and without infractions. A balanced training set is what we used to avoid biassing our data and, thus, the models we built.

According to the database, 66 percent of instances in the test set had violations, which matches the test set's sample size. Among the 66 labels in the training set, 45 are absent, while another 11 appear in less than half of the instances. As a result, the data in this research may be used to assess the effectiveness of few-shot learning.

[5]

**Targets of Legal Prediction: (Existing Problem)**

1. **Multi-label Violation**

One of the primary responsibilities is to determine whether human rights articles and/or procedures have been violated (if any). Up to this point, the European Convention on Human Rights has 66 articles and protocols in total. We use a multi-label classification job to do this, and when there is no violation, no labels are issued [6].

1. **Case Classification:**

The relevance of a case is predicted on a scale from 1 (key case) to 4 (insignificant). The European Court of Human Rights (ECHR) provides legal practitioners with these ratings, which indicate a case's contribution to the evolution of case law. Nearly three-quarters of all papers in the dataset received a score of 1, which indicates that only a small percentage of them merit further study; the rest were given values of 2 (904), 3 (2982), and 4 (6,496). This mean’s only 10% are useful cases, while remaining are useless and never be used for further studies [6].

1. **Binary Violation:**

Human rights articles and protocols that have been broken should be classified as "positive," and those that haven't should be considered "negative."

**Experiments Conducted:**

1. **Textual analysis:**

Researchers utilized tenfold cross-validation to check that the model performed well in general and was not unduly sensitive to the collection of instances on which it was trained. Using tenfold cross-validation instead of threefold cross-validation provides additional data for training (i.e., 90 percent rather than 66.7 percent).   the findings under ‘cross-val'.

Also note that the number of ‘violation' instances equals the number of ‘no violation' cases, since we utilized a balanced dataset (both for cv and t).

So, if we merely estimate the result at random, we'd be right approximately half the time. Percentages over 50% suggest that the model can employ (simplified) textual information to enhance case prediction. Table 5 presents the cross-validation findings first.

To assess the model's performance, researchers employ accuracy, recall, and f-score. Precision is the proportion of situations when the label (‘violation' or ‘no violation') is right. Recall is the proportion of properly labelled instances. The F-score is the mean of accuracy and recall [7].

1. **Future Prediction Analysis:**

It is observed in many experiment of future prediction that random selection of cases is easier to train on, training and predicting for two separate times. Furthermore, the quantity of training data has no discernible impact on the outcomes. In Experiment 1 and Experiment 1\*, the selected articles averaged 0.77, which is almost identical to the results of Experiment 1. However, when tested on distinct times, the accuracy was much lower. Thus, anticipating future judgments may be more difficult than previously thought. This difficulty rises as the gulf between training and testing data widens [8].

1. **Judges Evaluation:**

Only the outcome of the case was considered in this experiment, not how each judge voted. If judges have a reputation for ruling in favor of breaches (or non-violations) of the European Convention on Human Rights (ECHR), this does not always translate into a preference for violating that specific ECHR article. Regardless of the judge's personal view, he or she is more likely to be in a chamber that votes in favor of a violation. Judges are given varying weights based on the kind of material they are evaluating [8].

**Future Work and Limitation:**

No rationale exists for the predictions of neural models, which surpass prior feature-based models. Legal practitioners should not rely just on attention scores to determine which elements of a book have the most impact on their predictions; see [9]. In the NLP community, proper reasons are an essential issue that has to be addressed in the future. It is our hope that expanding the scope of this research to include automated analysis of other resources (such as applicable case law, dockets and earlier judgements) would allow us to enhance system performance and provide more support for system choices made using multiple inputs. The European Court of Justice, the US Supreme Court, and many languages will also be used to test neural approaches for their ability to anticipate judicial outcomes. Once these models have been adapted, we want to investigate multitask learning and apply it to data from different courts [10]

# References

|  |  |
| --- | --- |
| [1] | Ramesh Nallapati and Christopher D. Manning Stumbles, "Legal docket classification: Where machine learning," in *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, Honolulu, Hawaii, 2008. |
| [2] | Zhong, Haoxi and, "Legal Judgment Prediction via Topological Learning," in *Association for Computational Linguistics*, Brussels, Belgium, 2018. |
| [3] | Finale Doshi-Velez, Been Kim, "Towards A Rigorous Science of Interpretable Machine Learning," *arXiv:1702.08608,* 2017. |
| [4] | Aletras N, Tsarapatsanis D, Preoţiuc-Pietro D, Lampos V, "Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspective," *PeerJ Computer Science 2:e93,* p. 11, 2016. |
| [5] | "internet Archieves," Wednesday feb 2022. [Online]. Available: https://archive.org/details/ECHR-ACL2019.. [Accessed 13-feb-2022 feb 2022]. |
| [6] | Ilias Chalkidis, Ion Androutsopoulos, Nikolaos Aletras, "Neural Legal Judgment Prediction in English," *Computation and Language,* p. 7, 29. |
| [7] | Aletras N, Tsarapatsanis D, Preoţiuc-Pietro D, Lampos V., "Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspective.," *PeerJ Computer Science 2:e93,* p. 10, 2016. |
| [8] | Masha Medvedeva1,2 · Michel Vols2 · Martijn Wieling1, "Using machine learning to predict decisions of the European Court of Human Rights," *Artificial Intelligence and Law (2020,* p. 30, 2019. |
| [9] | Sarthak Jain and Byron C. Wallace, "Attention is not Explanation," in *Association for Computational Linguistics*, Minneapolis, Minnesota, 2019. |
| [10] | Zikun Hu, Xiang Li, Cunchao Tu, Zhiyuan Liu, and Maosong Sun, "Few-Shot Charge Prediction with Discriminative Legal Attributes," in *27th International Conference on Computational Linguistics*, New York, 2018. |